MADCAP — The Microwave Anisotropy Dataset Computational Analysis Package

Julian Borrill

Scientific Computing Group,
National Energy Research Scientific Computing Center,
Lawrence Berkeley National Laboratory
&
Center for Particle Astrophysics,
University of California at Berkeley

Abstract

In the standard model of cosmology the universe starts with a hot Big Bang. As the universe expands it cools, and after 300,000 years it drops below the ionisation temperature of hydrogen. The previously free electrons become bound to protons, and with no electrons for the photons to scatter off they continue undeflected to us today. This image of the surface of last-scattering is what we call the Cosmic (because it fills the universe) Microwave (because of the frequency at which its black body spectrum peaks today) Background (because it originates behind all other light sources). Despite its stunning uniformity - isotropic to a few parts in a million - it is the tiny perturbations in the CMB that give us an unprecedented view of the early universe. First detected by the COBE satellite in 1991, these anisotropies are an imprint of the primordial density fluctuations needed to seed the development of gravitationally bound objects in the universe, and are potentially the most powerful discriminant between cosmological models.

Realizing the extraordinary scientific potential of the CMB requires precise measurements of these tiny anisotropies over a significant fraction of the sky at very high resolution. The analysis of the resulting datasets is a serious computational challenge. Existing algorithms require terabytes of memory and hundreds of years of CPU time. We must therefore both maximize our resources by moving to supercomputers and minimize our requirements by algorithmic development. Here we will outline the nature of the challenge, present our current optimal algorithm, discuss its implementation - as the MADCAP software package - and its application to data from the North American test flight of the joint Italian-U.S. BOOMERanG experiment on the Cray T3E at NERSC and CINECA.

A documented β-release of MADCAP is publicly available at

http://cfpa.berkeley.edu/~borrill/cmb/madcap.html

1 Introduction

The current standard model of cosmology starts with a hot Big Bang. Whilst there is still debate about what this actually means, it is generally agreed is that what emerges is a very hot, expanding, space-time — the Universe. As the Universe expands its temperature falls and about 300,000 years after the Big Bang it cools to below the ionisation temperature of hydrogen and the previously free electrons become bound to protons. With no electrons to scatter off, the photons propagate undeflected through space to the present. When we observe this radiation today we are seeing the universe as it was when it was 1/40,000th of its present age. This image of the epoch when the primordial photons last scattered is what we call the Cosmic Microwave Background (CMB) radiation.

Detected serendipitously in 1969, the CMB was considered the decisive argument in the debate of the day between the 'Steady State' and 'Big Bang' cosmologies. Today, extraordinary instruments are measuring the tiniest variations in the CMB photons' temperature in different directions, and the results they are giving hold the promise of settling this generation's cosmological debates [1]. They have the potential to describe the geometry of space-time, showing whether straight lines continue forever or not; they can tell us how much mass and energy the universe contains, and the forms that it takes; and they may shed light the kinds of things that could have happened in the very first moments of the Big Bang.

As observers began to search for fluctuations in the CMB temperature they soon discovered that it was astonishingly uniform – 2.735 Kelvin in every direction. Building ever more sensitive detectors, the first variations weren't discovered until the milliKelvin regime. However these were due our galaxy's motion through space, a Doppler effect making the universe appear hotter in the direction in which we are going and colder in the direction from which we have come. Finally in 1992 the COBE satellite team reported intrinsic spatial variation in the CMB's temperature of around $\pm 30\mu K$ when averaged over patches of sky approximately 10^o across.

Since that detection the focus has been on measuring the extent of the variation with ever greater precision on ever smaller angular scales. Both the absolute and the relative power of the fluctuations on different angular scales contain a wealth of cosmological information. For example, we expect to see a peak in the power at the angular scale corresponding to the horizon size (the limiting scale for coherent physical processes) at last scattering. The apparent angular size on the sky today of a particular physical scale in the past depends on the geometry of the Universe – decreasing in size as the curvature of space increases. A measurement of the location of the peak in the angular power spectrum therefore tells us whether the Universe is open, flat or closed. Similarly the height of this peak is related to the total energy density in the universe, and how it is distributed. The presence or absence of lower secondary peaks, at resonances of the fundamental scale, may allow us to rule out some classes of theories of the origins of the first density perturbations in the universe.

Making an observation with small enough statistical and systematic error bars to resolve the detailed shape of this angular power spectrum requires very sensitive detectors scanning a significant fraction of the sky at very high resolution. These conditions are now being met for the first time in balloon-borne experiments such the joint Italian–U.S. BOOMERanG project. Lifted into the stratosphere to minimize atmospheric interference, this experiment measures changes in the voltage across an extremely sensitive bolometer, cooled to a few milliKelvin, as it scans the sky. Since the voltage across a bolometer depends on it's temperature, hidden within such a data set is a measurement of CMB temperature fluctuations.

Reducing tens of millions of individual observations to an angular power spectrum of a thousand multipoles is a computationally challenging task. Each observation contains detector noise as well as signal on the sky; the signal on the sky includes not only the CMB but also foreground sources of microwave radiation such as interstellar dust; and both the noise and the signal components of nearby observations are correlated. At present, except in very restrictive circumstances, algorithms to solve the equations relating the time-ordered data to a map of the sky temperature, and then to the angular power spectrum, scale as the number of map pixels squared in memory and cubed in floating point operation count. The COBE map contained only a few thousand pixels, but current balloon observations are generating maps with tens and hundreds of thousands pixels, and the MAP and PLANCK satellites — to be launched in 2000 and 2007 respectively — will increase this to millions and tens of millions.

To realize the full potential of these CMB observations we simultaneously need to maximize our computational resources (by moving to supercomputers) and to minimize our computational requirements (by optimizing our algorithms and implementations). Here we present our current optimal algorithm and discuss its implementation in the MADCAP software package and describe its application at NERSC and CINECA to data from the North American test flight of BOOMERanG. For simplicity we will consider an observation from a single detector with no significant foreground contamination — a situation that was realized in practice in the analysis of the best channel of the BOOMERanG North America data.

In what follows vectors and matrices are written in plain fonts in the time domain and italic fonts in the pixel domain.

2 From The Time-Ordered Data To The Map

2.1 Algorithm

Our first step is to translate the observation from the temporal to the spatial domain — to make a map [2]. Knowing where the detector was pointing, (θ_t, ψ_t), at each of the $\mathcal{N}_t$ observation, and dividing the sky into $\mathcal{N}_p$ pixels, we can construct an $\mathcal{N}_t \times \mathcal{N}_p$ pointing matrix A whose entries give the weight of pixel p in observation t. For a total power scanning experiment such as BOOMERanG this has a particularly simple form

$$\mathrm{A}_{tp} = \begin{cases} 1 & \text{if} \quad (\theta_t, \psi_t) \in p \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

while a more complex observing strategy would give a correspondingly complex structure.

The time-ordered data vector can now be written

$$\mathrm{d} = \mathrm{A}\, s + \mathrm{n} \tag{2}$$

in terms of the pixelised CMB signal s and time-stream noise n. Under the assumption of Gaussianity, the noise probability distribution is

$$\mathrm{P}(\mathrm{n}) = (2\pi)^{-\mathcal{N}_t/2} \exp\left\{ -\frac{1}{2} \left(\mathrm{n}^{\mathrm{T}}\, \mathrm{N}^{-1}\, \mathrm{n} + \mathrm{Tr}\left[\ln \mathrm{N}\right] \right) \right\} \tag{3}$$

where N is the time-time noise correlation matrix given by

$$\mathrm{N} \equiv \langle \mathrm{n}\, \mathrm{n}^{\mathrm{T}} \rangle \tag{4}$$

We can now use equation (2) to substitute for the noise in equation (3), so the probability that, with a particular underlying CMB signal, we would have obtained the observed time-ordered data is

$$\mathrm{P}(\mathrm{d}|s) = (2\pi)^{-\mathcal{N}_t/2} \exp\left\{ -\frac{1}{2} \left((\mathrm{d} - \mathrm{A}\, s)^{\mathrm{T}}\, \mathrm{N}^{-1}\, (\mathrm{d} - \mathrm{A}\, s) + \mathrm{Tr}\left[\ln \mathrm{N}\right] \right) \right\} \tag{5}$$

Assuming that all CMB signals are *a priori* equally likely, this is proportional to the likelihood of the CMB signal given the time-ordered data. Maximizing over s now gives the maximum likelihood pixelized data (or map) d

$$d = \left(\mathrm{A}^{\mathrm{T}}\, \mathrm{N}^{-1}\, \mathrm{A} \right)^{-1} \mathrm{A}^{\mathrm{T}}\, \mathrm{N}^{-1}\, \mathrm{d} \tag{6}$$

Substituting back for the time-ordered data in equation (6) we recover the obvious fact that this pixelized data is the sum of the true CMB signal and some residual pixelized noise

$$\begin{aligned} d &= \left(\mathrm{A}^{\mathrm{T}}\, \mathrm{N}^{-1}\, \mathrm{A} \right)^{-1} \mathrm{A}^{\mathrm{T}}\, \mathrm{N}^{-1}\, (\mathrm{A}\, s + \mathrm{n}) \\ &= s + n \end{aligned} \tag{7}$$

where this pixel noise

$$n = \left(\mathrm{A}^{\mathrm{T}}\, \mathrm{N}^{-1}\, \mathrm{A} \right)^{-1}\, \mathrm{A}^{\mathrm{T}}\, \mathrm{N}^{-1}\, \mathrm{n} \tag{8}$$

has correlations given by

$$\begin{aligned} N &= \langle n\, n^{\mathrm{T}} \rangle \\ &= \left(\mathrm{A}^{\mathrm{T}}\, \mathrm{N}^{-1}\, \mathrm{A} \right)^{-1} \end{aligned} \tag{9}$$

The map-making algorithm can therefore be divided into two steps,

M1 – construct the inverse pixel-pixel noise correlation matrix and noise-weighted map

$$N^{-1} = \mathrm{A}^{\mathrm{T}}\, \mathrm{N}^{-1}\, \mathrm{A}$$

$$z \equiv N^{-1}\, d = \mathrm{A}^{\mathrm{T}}\, \mathrm{N}^{-1}\, \mathrm{d}$$

M2 – solve for the pixelized data

$$d = (N^{-1})^{-1}\, z$$

which are encoded in the two map-making modules — *inv_pp_noise.c* and *p_data.c* — in MADCAP.

2.2 Implementation

The first half of Table 1 shows the computational cost of a brute force implementation of each of the steps in the map-making algorithm (recall that multiplying an $[a \times b]$ matrix and a $[b \times c]$ matrix in general requires $2\,a\,b\,c$ operations). For the BOOMERanG North America data — with $\mathcal{N}_t \sim 1.5 \times 10^6$ and $\mathcal{N}_p \sim 2.4 \times 10^4$ — simply making the map would require 9 Tb disc space (storing data in 4-byte precision), 18 Tb RAM[1] (doing all calculations in 8-byte precision) and 10^{17} floating point operations.

Calculation	Brute Force			Structure-Exploiting		
	Disc	RAM	Flops	Disc	RAM	Flops
M1	$4\,\mathcal{N}_t^2$	$8\,\mathcal{N}_t^2$	$2\,\mathcal{N}_t^2\,\mathcal{N}_p$	$4\,(\mathcal{N}_p^2 + \mathcal{N}_t)$	$8\,(\mathcal{N}_p^2 + \mathcal{N}_t)$	$3\mathcal{N}_\tau\,\mathcal{N}_t$
M2	$4\,\mathcal{N}_p^2$	$8\,\mathcal{N}_p^2$	$(2 + \frac{2}{3})\,\mathcal{N}_p^3$	$4\mathcal{N}_p^2$	$8\,\mathcal{N}_p^2$	$(2 + \frac{2}{3})\,\mathcal{N}_p^3$
Total	$4\,\mathcal{N}_t^2$	$8\,\mathcal{N}_t^2$	$2\,\mathcal{N}_t^2\,\mathcal{N}_p$	$4\,(\mathcal{N}_p^2 + \mathcal{N}_t)$	$8\,(\mathcal{N}_p^2 + \mathcal{N}_t)$	$(2 + \frac{2}{3})\,\mathcal{N}_p^3$

Table 1: Computational requirements for the map-making algorithm

Fortunately there are two crucial structural features to be exploited here. As noted above, for a simple scanning experiment like BOOMERanG the pointing matrix A is very sparse, with only a single 1 in each row. Moreover, the inverse time-time noise correlations are (by fiat) both stationary and fall to zero beyond some time-separation much shorter than the duration of the observation

$$\begin{aligned} N_{tt'}^{-1} &= f(|t - t'|) \\ &= 0 \qquad \forall \quad |t - t'| > \tau \ll \mathcal{N}_t \end{aligned} \tag{10}$$

so that the inverse time-time noise correlation matrix is symmetric and band-diagonal, with bandwidth $\mathcal{N}_\tau = 2\tau + 1$. The second half of Table 1 shows the impact of exploiting this structure on the cost of each step. The limiting

[1] Since all algorithms here will be operation-count limited we always assume in-core implementations. Out-of-core methods would reduce memory requirements but prohibitively increase the run-time overhead, for example in the cache hit cost of moving from level 3 (matrix-matrix) to level 1 (vector-vector) BLAS.

step is now no longer constructing the inverse pixel-pixel noise correlation matrix but inverting it and solving for the map. Although this inversion is not necessary to obtain just the map, the current power spectrum algorithm requires the pixel-pixel noise correlation matrix itself and not simply its inverse. We therefore Cholesky decompose the positive definite matrix N^{-1}, and use the decomposition both to solve for the map and to calculate N. For the same dataset making the map now takes 2.3 Gb of disc, 4.6 Gb of RAM, and 3.7×10^{13} flops.

Although the final analysis of the data is dominated by the second step, 'quick and dirty' systematics tests can be performed at much lower map resolution (ie. with much larger pixels) to reduce $\mathcal{N}_p$ by up to an order of magnitude. At this point it becomes important to optimize the first step too. The structure exploiting algorithm reduces these calculations to:

for each observation (at time t, of pixel p)
    for each observation within τ of it (at time t', of pixel p')
        add $f(|t-t'|)$ to $N^{-1}_{pp'}$
        add $f(|t-t'|)\, d_{t'}$ to z_p

Dividing this into blocks of contiguous p-pixels allows each processor to work independently. However the number of times a pixel is observed can vary from a few to thousands, so a simple division into equal numbers of p-pixels on each processor can be very poorly load-balanced. The MADCAP implementation therefore starts by determining the number of operations required for each p-pixel and dividing them among the processors in blocks which are still contiguous but whose numbers of p-pixels vary so that the total operation count in each block is as close to the mean count per processor as possible. Although this means that the memory requirement per processor varies (and is not known exactly in advance) for the BOOMERanG North America data it reduced the run-time of these steps by a factor of two or more.

3 From The Map To The Power Spectrum

3.1 Algorithm

We now want to move to a basis where the CMB observation can be compared with the predictions of various cosmological theories — the angular power spectrum. We decompose the CMB signal at each pixel in spherical harmonics

$$s_p = \sum_{lm} a_{lm}\, B_l\, Y_{lm}(\theta_p, \psi_p) \tag{11}$$

where B is the pattern of the observation beam (assumed to be circularly symmetric) in l-space. The correlations between such signals then become

$$S_{pp'} \equiv \langle s_p\, s_{p'} \rangle = \sum_{lm} \sum_{l'm'} \langle a_{lm}\, a_{l'm'} \rangle\, B_l\, B_{l'}\, Y_{lm}(\theta_p, \psi_p)\, Y_{l'm'}(\theta_{p'}, \psi_{p'}) \tag{12}$$

For isotropic fluctuations the correlations depend only on the angular separation

$$\langle a_{lm}\, a_{l'm'} \rangle = C_l\, \delta_{ll'}\, \delta_{mm'} \tag{13}$$

and the pixel-pixel signal correlation matrix entries become

$$S_{pp'} = \sum_{l} \frac{2l+1}{4\pi}\, B_l^2\, C_l\, P_l(\chi_{pp'}) \tag{14}$$

where P_l is the Legendre polynomial and $\chi_{pp'}$ the angle between the pixel pair p, p'. These C_l multipole powers completely characterize a Gaussian CMB, and are an otherwise model-independent basis in which to compare theory with observations. The finite beam size means that any experiment has a maximum multipole sensitivity, above which all power is beam-smeared. Coupled with incomplete sky coverage, this means that in practice the C_l

that we extract do *not* form a complete orthonormal basis. We therefore group the $\mathcal{N}_l$ accessible multipoles into $\mathcal{N}_b$ bins, adopting a fixed spectral shape function C_l^s and characterizing the CMB signal by its bin powers C_b with

$$C_l = C_{b:l\in b}\, C_l^s \tag{15}$$

Since the signal and noise are assumed to be realizations of independent Gaussian processes the pixel-pixel map correlations are

$$\begin{aligned} D &\equiv \langle d\, d^{\mathrm{T}} \rangle \\ &= \langle s\, s^{\mathrm{T}} \rangle + \langle n\, n^{\mathrm{T}} \rangle \\ &= S + N \end{aligned} \tag{16}$$

and the probability distribution of the map given a particular power spectrum C is now

$$P(d|C) = (2\pi)^{-\mathcal{N}_p/2} \exp\left\{ -\frac{1}{2} \left(d^{\mathrm{T}}\, D^{-1}\, d + \mathrm{Tr}\left[\ln D\right] \right) \right\} \tag{17}$$

Assuming a uniform prior for the spectra, this is proportional to the likelihood of the power spectrum given the map. Maximizing this over C then gives us the required result, namely the most likely CMB power spectrum underlying the original observation d.

Finding the maximum of the likelihood function of equation (17) is a much harder problem than making the map. Since there is no closed-form solution corresponding to equation (6) we must find both a fast way to evaluate the likelihood function at a point, and an efficient way to search the $\mathcal{N}_b$-dimensional parameter space for the peak. The fastest general method extant is to use Newton-Raphson iteration to find the zero of the derivative of the logarithm of the likelihood function [3]. If the log likelihood function

$$\mathcal{L}(C) = -\frac{1}{2} \left(d^{\mathrm{T}}\, D^{-1}\, d + \mathrm{Tr}\left[\ln D\right] \right) \tag{18}$$

were quadratic, then starting from some initial guess at the maximum likelihood power spectrum C_o the correction $\delta\mathcal{C}_o$ that would take us to the true peak would simply be

$$\delta C_o = -\left(\left[\frac{\partial^2 \mathcal{L}}{\partial C^2} \right]^{-1} \frac{\partial \mathcal{L}}{\partial C} \right)_{C=C_o} \tag{19}$$

Since the log likelihood function is not quadratic, we now take

$$C_1 = C_o + \delta C_o \tag{20}$$

and iterate until $\delta C_n \sim 0$ to the desired accuracy. Because any function is approximately quadratic near a peak, if we start searching sufficiently close to a peak this algorithm will converge to it. Of course there is no guarantee that it will be the global maximum, and in general there is no certainty about what 'sufficiently close' means in practice. However experience to date suggests that the log likelihood function is sufficiently strongly singly peaked to allow us to use this algorithm with some confidence.

The core of the algorithm is then to calculate the first two derivatives of the log likelihood function with respect to the multipole bin powers

$$\frac{\partial \mathcal{L}}{\partial C_b} = \frac{1}{2} \left(d^{\mathrm{T}}\, D^{-1}\, \frac{\partial S}{\partial C_b}\, D^{-1}\, d - \mathrm{Tr}\left[D^{-1}\, \frac{\partial S}{\partial C_b} \right] \right) \tag{21}$$

$$\frac{\partial^2 \mathcal{L}}{\partial C_b \partial C_{b'}} = -d^{\mathrm{T}}\, D^{-1}\, \frac{\partial S}{\partial C_b}\, D^{-1}\, \frac{\partial S}{\partial C_{b'}}\, D^{-1}\, d + \frac{1}{2} \mathrm{Tr}\left[D^{-1}\, \frac{\partial S}{\partial C_b}\, D^{-1}\, \frac{\partial S}{\partial C_{b'}} \right] \tag{22}$$

Each iteration of the power-spectrum extraction algorithm can therefore be divided into six steps,

P1 – Calculate the $\mathcal{N}_b$ pixel-pixel signal correlation bin derivative matrices

$$\frac{\partial S}{\partial C_b} = \sum_{l \in b} \frac{2l+1}{4\pi} B_l^2 \, C_l^s \, P_l(\chi_{pp'})$$

P2 – Construct the pixel-pixel map correlation matrix, Cholesky decompose it, and triangular solve for the data-weighted map

$$D = N + \sum_b C_b \frac{\partial S}{\partial C_b} = L \, L^{\mathrm{T}} \quad \& \quad L \, L^{\mathrm{T}} \, z = d$$

P3 – Triangular solve the linear systems

$$L \, L^{\mathrm{T}} \, W_b = \frac{\partial S}{\partial C_b}$$

P4 – Assemble the first derivative

$$\frac{\partial \mathcal{L}}{\partial C_b} = \frac{1}{2} \left(d^{\mathrm{T}} \, W_b \, z - \mathrm{Tr} \, [W_b] \right)$$

P5 – Assemble the second derivative

$$\frac{\partial^2 \mathcal{L}}{\partial C_b \partial C_{b'}} = -d^{\mathrm{T}} \, W_b \, W_{b'} \, z + \frac{1}{2} \mathrm{Tr} \, [W_b \, W_{b'}]$$

P6 – Calculate the spectral correction

$$\delta C = - \left[\frac{\partial^2 \mathcal{L}}{\partial C^2} \right]^{-1} \frac{\partial \mathcal{L}}{\partial C}$$

which are encoded in the six power-spectrum extraction modules — *pp_signal.c*, *L.c*, *trisolve.c*, *dLdC.c*, *d2LdC2.c* and *dC.c* — in MADCAP.

3.2 Implementation

Table 2 shows the computational cost of each of the steps in each iteration of the power-spectrum algorithm. Obtaining the maximum likelihood power spectrum for the BOOMERanG North America data — with $\mathcal{N}_p \sim 2.4 \times 10^4$, $\mathcal{N}_b \sim 10$ and $\mathcal{N}_l \sim 1200$ — requires 50 Gb of disc space, 9.2 Gb of RAM and 2.8×10^{14} flops.

Calculation	Disc	RAM	Flops
P1	$4 \mathcal{N}_b \mathcal{N}_p^2$	$16 \mathcal{N}_p^2$	$O(\mathcal{N}_l \mathcal{N}_p^2)$
P2	$4 \, (\mathcal{N}_b + 2) \, \mathcal{N}_p^2$	$16 \mathcal{N}_p^2$	$\frac{2}{3} \mathcal{N}_p^3$
P3	$4 \, (2 \mathcal{N}_b + 1) \, \mathcal{N}_p^2$	$16 \mathcal{N}_p^2$	$2 \mathcal{N}_b \mathcal{N}_p^3$
P4	$4 \mathcal{N}_b \mathcal{N}_p^2$	$8 \mathcal{N}_p^2$	$2 \mathcal{N}_b \mathcal{N}_p^2$
P5	$4 \mathcal{N}_b \mathcal{N}_p^2$	$16 \mathcal{N}_p^2$	$3 \mathcal{N}_b^2 \mathcal{N}_p^2$
P6	$4 \mathcal{N}_b^2$	$8 \mathcal{N}_b^2$	$(2 + \frac{2}{3}) \, \mathcal{N}_b^3$
Total	$4 \, (2 \mathcal{N}_b + 2) \, \mathcal{N}_p^2$	$16 \mathcal{N}_p^2$	$(2 \mathcal{N}_b + \frac{2}{3}) \, \mathcal{N}_p^3$

Table 2: Computational requirements for each iteration of the power spectrum algorithm

Note that in step P1 we calculate the pixel-pixel signal correlation bin derivative matrices (which we need in step P3 anyway), rather than the full pixel-pixel signal correlation matrix. Since these are independent of the bin power step P1 does not need to be repeated at each iteration. The trade-off here is the near doubling of disc space required, since we now want to keep these matrices throughout the calculation, and not overwrite them in step P3. If disc space is at a premium, however, we can simply revert to recalculation. In step P2 we calculate all the terms necessary to construct the log-likelihood itself (18), which is therefore generated as a useful by-product, allowing us to confirm that the algorithm is moving to higher likelihood and to check the appropriateness of our convergence criterion. In practice over 80% of the run time is spent in the triangular solves of step P3. Since these involve level 3 BLAS only they are highly optimized, and we typically reach 40–80 % peak performance on the T3E both at CINECA and at NERSC, with the fraction slowly decreasing with the total number of processors used

4 Conclusions

MADCAP is a highly optimized, portable, parallel implementation (using ANSI C, MPI and the ScaLAPACK libraries) of the current optimal general algorithm for extracting the most useful cosmological information from total-power observations of the CMB. The β-release of MADCAP has been ported to a wide range of parallel platforms — including the Cray T3E, SGI Origin 2000, HP Exemplar and IBM SP2. The combination of parallelism and algorithmic optimization has enabled CMB datasets that would previously have been intractable to be analyzed in a matter of hours. We have successfully applied MADCAP at CINECA and NERSC to the data from the North American Test flight of the BOOMERanG experiment; the scientific results will be published shortly [4].

Existing algorithms are capable of dealing with CMB datasets with up to 10^5 pixels. Over the next 10 years a range of observations are expected to produce datasets of 5×10^5 (BOOMERanG LDB), 10^6 (MAP) and 10^7 (PLANCK) pixels. As shown in Table 3 (where we have assumed a constant 10 multipole bins and 5 Newton-Raphson iterations) the scaling with projected map size pushes the analysis of these observations well beyond the capacity of even the most powerful current supercomputers.

Flight	$\mathcal{N}_p$	Disc	RAM	Flops
BOOMERanG NA	24,000	50 Gb	9 Gb	1.4×10^{15}
MAXIMA 1	32,000	90 Gb	16 Gb	3.2×10^{15}
MAXIMA 2	80,000	560 Gb	100 Gb	5.1×10^{16}
BOOMERanG LDB	450,000	18 Tb	3 Tb	3.7×10^{18}
PLANCK	20,000,000	35 Pb	2.4 Pb	8×10^{23}

Table 3: MADCAP computational requirements for current & forthcoming CMB observations

Such datasets will require new algorithms. The limiting steps in the above analysis are associated with operations involving the pixel-pixel correlation matrices for the noise N, the signal S, and most particularly their sum D. The problem here is the noise and the signal have different natural bases. The inverse noise correlations are symmetric, band-diagonal and approximately circulant in the time domain, while the signal correlations are diagonal in the spherical harmonic domain. Moreover, thus far we have ignored foreground sources which will be spatially correlated and so most naturally expressed in the pixel domain. The fundamental problem is then that the data correlations — being the sum of these three terms — are complex in all three domains. Developing the approximations necessary to handle such data sets is an area of ongoing research.

5 Acknowledgments

The author would like to thank Amadeo Balbi, Andrew Jaffe, Pedro Ferriera, Shaul Hanany and Radek Stompor of the COMBAT collaboration, Xiaoye Li and Osni Marques of the NERSC Scientific Computing Group, and Paolo de Bernardis, Andrew Lange, Silvia Masi, Phil Mauskopf, Barth Netterfield, John Ruhl and the other members of the BOOMERanG team.

References

[1] L. Knox, *Phys. Rev.* **D52**, 4307 (1995), astro-ph/9504054; C. Jungman, M. Kamionkowski, A. Kosowsky and D. Spergel, *Phys. Rev.* **D54**, 1332 (1996), astro-ph/9512139; W. Hu, N. Sugiyama and J. Silk, *Nature* **386**, 37 (1997), astro-ph/9604166; J. R. Bond, G. Efstathiou and M. Tegmark, *MNRAS* **291**, 33 (1997), astro-ph/9702100; M. Zaldarriaga, D. Spergel and U. Seljak, *Ap. J* **488**, 1 (1997), astro-ph/9702157.

[2] E. L. Wright, astro-ph/9612006; M. Tegmark, *Ap. J. Lett* **480**, 87 (1997), astro-ph/9611130; A. H. Jaffe, J. R. Bond, P. G. Ferreira and L. Knox in "Proceedings of the 3K Cosmology: EC-TMR Conference", American Institute of Physics, (1999).

[3] K. M. Górski, *Ap. J.* **430** L85 (1994), astro-ph/9403066; M. Tegmark, *Phys. Rev.* **D55**, 5895 (1997), astro-ph/9611174; J. R. Bond, A. H. Jaffe and L. Knox, *Phys. Rev.* **D57** 2117 (1998), astro-ph/9708203; J. Borrill, *Phys. Rev.* **D59**, 7302 (1999), astro-ph/9712121; J. Borrill in "Proceedings of the 3K Cosmology: EC-TMR Conference", American Institute of Physics, (1999), astro-ph/9903204.

[4] P. Mauskopf *et al*, A. Melchiorri *et al* in preparation.